PROGRAMMATION D'UN SYSTÈME DE GESTION DE BASE DE DONNÉES

ALGÈBRE RELATIONNELLE

Algèbre relationnelle

- Manipulations des relations par interrogation.
- Notation algébrique
- Base théorique du langage SQL.
- Combinaisons de relations avec différents opérateurs pour obtenir de nouvelles relations.

L'opérateur de projection

Définition

L'opérateur de **projection**, noté π, permet de ne retenir que les *n*-uplets des attributs indiqués par l'opérateur en supprimant les éventuels doublons (partition verticale).

Exemple

R

A	B	C	D
2	A	1	1
2	B	5	2
3	C	2	2
4	D	7	6
3	E	5	2
6	G	7	6

L'opérateur de sélection

Définition

L'opérateur de **sélection**, noté σ, permet de ne retenir que les *n*-uplets vérifiant une propriété particulière donnée sous la forme d'un prédicat (partition horizontale).

Exemples d'opérateurs : >, <, =, ≤, ≥, ≠, ⊂, ⊆, ⊄

Exemple

R

A	B	C	D
2	A	1	1
2	B	5	2
3	C	2	2
4	D	7	6
3	E	5	2
6	G	7	6

Exemple

Avion
N° série Type Capacité

Quel est le numéro de série des avions dont la capacité est supérieure à 150 passagers ?

L'opérateur de jointure

Définition

- Opération notée ⋈
- La jointure naturelle ou simplement jointure est le rapprochement de deux relations liées *via* des attributs communs dans leurs schémas correspondants.
- Les *n*-uplets du résultat sont obtenus par concaténation des attributs des deux relations lorsque les attributs communs ont des valeurs identiques (on simplifie en n'écrivant ces attributs qu'une seule fois).
- Équivalent à une sélection des valeurs de l'attribut commun sur le résultat d'un produit cartésien des relations concernées.

Autres opérateurs de jointure

- Θ-jointure : utilise d'autres opérateurs que l'égalité
- Jointure externe : récupération des n-uplets qui ne joignent pas.
- Semi-jointure droite ou gauche : jointure suivie d'une projection sur les attributs de la relation à droite ou à gauche de l'opérateur de jointure.

L'opérateur de jointure

Exemple

Nom de la personne qui consomme des crêpes ?

Consommation

N°P	NomP	Id
1	Crêpes	45
2	Cidre	89
5	Galettes	3

Clients

Id	NomC	Adresse
3	Parker	New-York
45	Simpson	Springfield
89	Kent	Metropolis

L'opérateur de jointure : l'auto-jointure

- Auto-jointure : jointure d'une relation avec elle-même.

N°Emp	Nom	N°Sup
1	Simpson	2
2	Kent	7
5	Kenobi	
7	Luthor	42

- Par rapport à l'exemple, à utiliser pour répondre à des questions du type « Qui est le supérieur hiérarchique de mon supérieur hiérarchique » ?
- Nécessité de renommer la relation pour identifier de façon unique chaque relation qui compose la jointure : opérateur ρ.

- $\rho_{S(A,\ B,\ C)}$(Employés) ⇒ Renomme la relation Employés(N°Emp, Nom, N°Sup) en S(A, B, C) le temps de la requête

L'opérateur de jointure : l'auto-jointure

- Jointure d'une relation avec elle-même.

N°Emp	Nom	N°Sup
1	Simpson	2
2	Kent	7
5	Kenobi	
7	Luthor	42

  - Quel est le supérieur du supérieur de Simpson ?

Opérations binaires

Soit deux relations **A(A_1, A_2, ..., A_n)** et **B(B_1, B_2, ..., B_n)**. **A** et **B** sont **compatibles** ssi elles ont le **même degré** et si **dom(A_i)=dom(B_i)** pour $1 \leq i \leq n$.

Définition

Ce sont les opérations mathématiques standards de la théorie des ensembles. Elles ne peuvent être appliquées que sur des relations compatibles.

Union : $A \cup B$. Relation qui inclut tous les *n*-uplets appartenant à A, à B ou au deux. Les doublons sont éliminés.

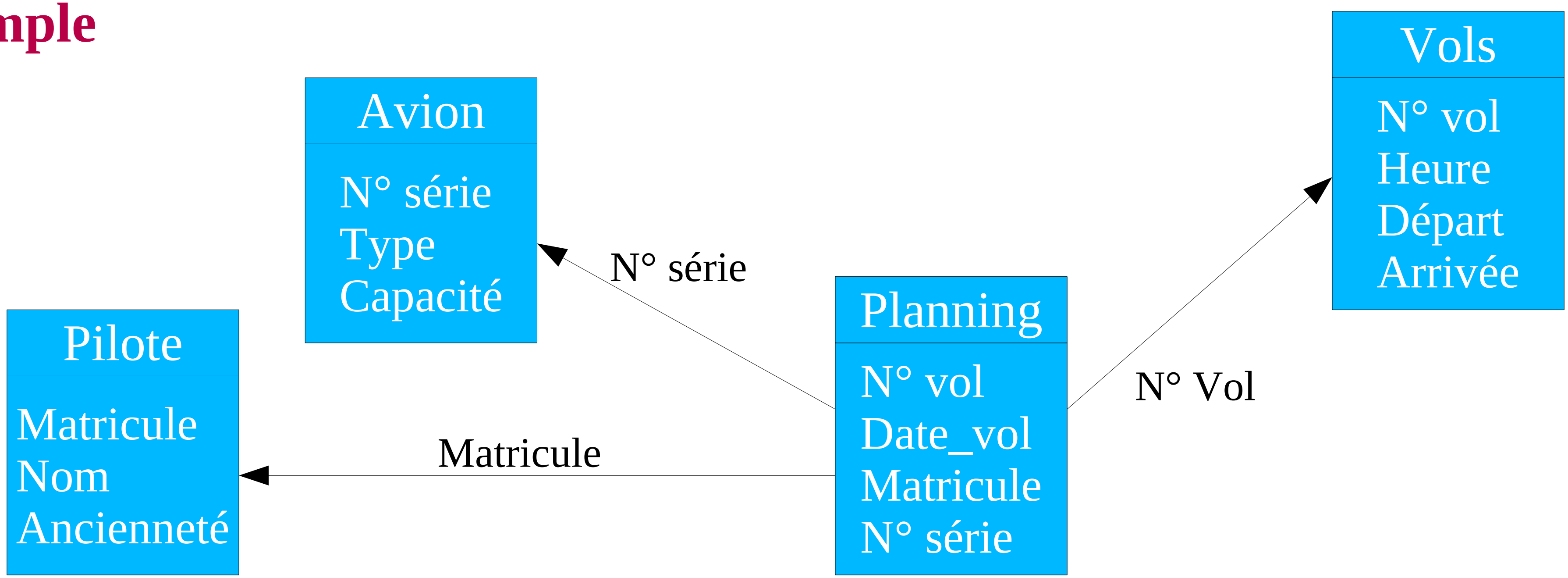
Intersection : $A \cap B$. Relation qui inclut tous les *n*-uplets appartenant à la fois à A et B.

Différence : A-B. Relation qui inclut tous les *n*-uplets appartenant à A mais pas à B.

La différence

Différence : A-B. Relation qui inclut tous les *n*-uplets appartenant à A mais pas à B.

Exemple



Quelles sont les lignes qui ne sont jamais parcourues par des "Airbus" ?

La division

Schéma de Q : tous les attributs de R n'appartenant pas à S

n-uplets de Q : *n*-uplets q_j de Q / $\forall$ *n*-uplets S_i de S, le *n*-uplet (q_j, s_j) est un *n*-uplet de R ⇒ $Q \times S \subseteq R$

Exemple

R

A	B
a1	b1
a2	b1
a1	b2

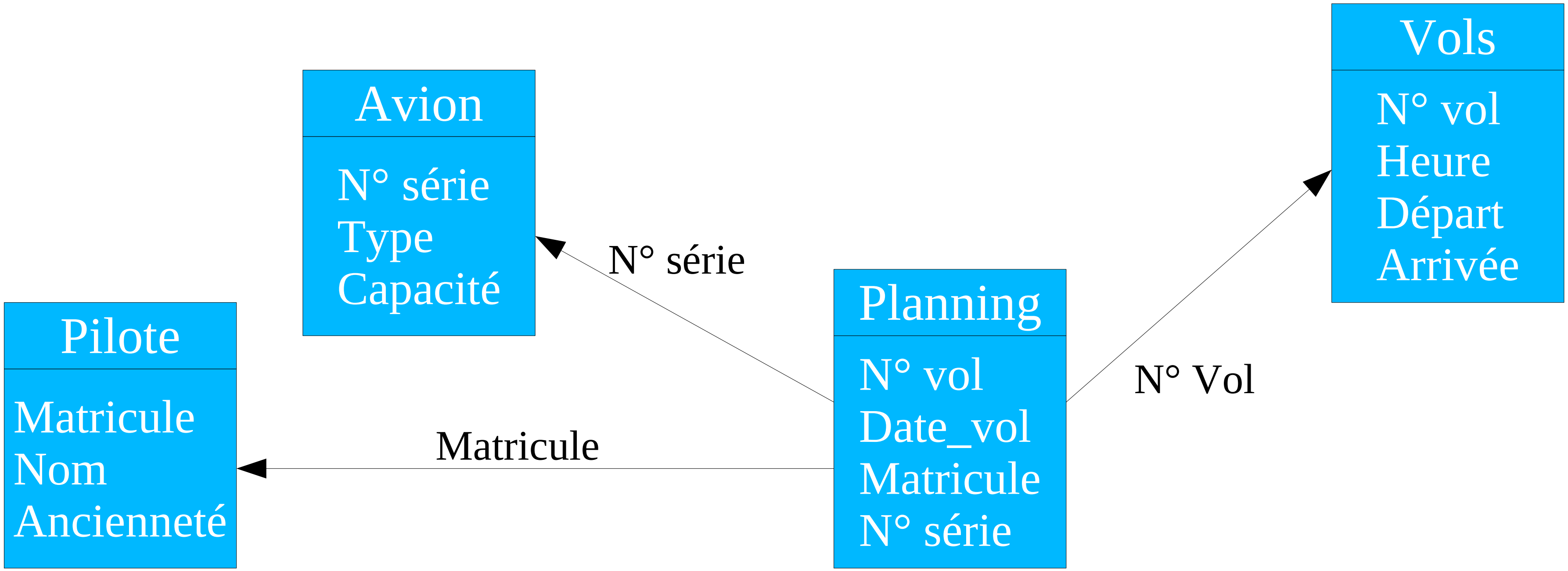
÷

S

B
b1
b2

=

La division



Quels sont les commandants qui volent sur tous les types d'avions ?

Calcul relationnel par n-uplet

- Les requêtes sont de la forme {t | P (t)}. Notation logique
- C’est l’ensemble des n-uplets tels que le prédicat P (t) est vrai pour t
- t est une variable n-uplet et t[A] désigne la valeur de l’attribut A dans t
- t ∈ r signifie que t est un n-uplet de r
- P est une formule de la logique de premier ordre

Rappel sur le calcul des prédicats

- Des ensembles d'attributs, de constantes, de comparateurs $\{<, \ldots\}$
- Les connecteurs logiques 'et' $\wedge$, 'ou' $\vee$ et la négation $\neg$
- Les quantificateurs $\exists$ et $\forall$:
  - $\exists t \in r\,(Q(t))$: Il existe un tuple t de r tel que Q est vrai
  - $\forall t \in r(Q(t))$: Q est vrai pour tout t de r

Exemples de requêtes

- Considérons les schémas de relations suivants :
  - Film(Titre, Réalisateur, Acteur) instance f
  - Programme(NomCiné, Titre, Horaire) instance p
  - f contient des infos sur tous les films et p concerne le programme à Bordeaux
- *Les films réalisés par Bergman*

$$\{t \mid t \in f \wedge t[\text{Réalisateur}] = \text{"Bergman"}\}$$

- *Les films où Jugnot et Lhermite jouent ensembles*

$$\{t \mid t \in f \wedge \exists s \in f\ (t[\text{Titre}] = s[\text{Titre}] \wedge t[\text{Acteur}] = \text{"Jugnot"} \wedge s[\text{Acteur}] = \text{"Lhermite"})\}$$

Exemples de requêtes

- *Les titres des films programmés à Bordeaux :*

  $$\{t \mid \exists s \in p\ (t[Titre] = s[Titre])\}$$

- *Les films programmés à l'UGC mais pas au Trianon :*

  $\{t \mid \exists s \in p\ (s[Titre] = t[Titre] \land s[NomCiné] = \text{« UGC »} \land \neg\exists u \in p(\ u[NomCiné] = \text{« Trianon »} \land u[Titre] = t[Titre]))\}$

- *Les titres de films qui passent à l'UGC ainsi que leurs réalisateurs :*

  $\{t \mid \exists s \in p\ (\exists u \in f\ (s[NomCiné] = \text{« UGC »} \land s[Titre] = u[Titre] = t[Titre] \land t[Réal] = u[Réal]))\}$

Retour sur la notion de clé

• Une clé est nécessaire pour identifier les *n*-uplets de façon unique sans en donner toutes les valeurs et respecter leur unicité

Définition

- Groupe minimum d'attributs qui détermine chaque *n*-uplet de façon unique.
- Plus formellement avec une relation R(U) et U l'ensemble des attributs :

$$X \text{ clé de } R(U) \text{ avec } X \subseteq U \text{ ssi } \forall r : R(U), \forall t_1, t_2 \in r,$$

$$t_1[X] = t_2[X] \Rightarrow t_1 = t_2$$

$$\text{et}$$

$$\nexists Y \subset X \Rightarrow t_1[Y] = t_2[Y] \Rightarrow t_1 = t_2$$

Expressions «saines »

Il est possible d'écrire des requêtes en calcul qui retournent une relation infinie.

Exemple : Soit NumCompte(Num) avec l'instance n et la requête $\{t \mid \neg t \in n\}$ i.e., les numéros de compte non recensés. Si on considère que le Dom(Num) = $\mathbb{N}$, alors la réponse à cette requête est infinie.

Une requête est **saine** si quelle que soit l'instance de la base dans laquelle on l'évalue, elle retourne une réponse finie.

Calcul relationnel par domaine

Les requêtes sont de la forme : $\left\{\left\langle x_1,\ldots,x_n\right\rangle \middle| P\left(x_1,\ldots,x_n\right)\right\}$

Les x_i représentent des variables de domaine (attributs de n-uplet).

$P\left(x_1,\ldots,x_n\right)$ est une formule similaire à celles que l'on trouve dans la logique des prédicats.

- Exemple : Les titres de films programmés à l'UGC de Bordeaux

$$\left\{\left\langle t\right\rangle \middle| \exists\left\langle nc,t,h\right\rangle \in p\left(nc=\mathrm{UGC}\right)\right\}$$

Relation entre les 3 langages

- Toute requête exprimée en algèbre peut être exprimée par le calcul.
- Toute requête “saine” du calcul peut être exprimée par une requête de l’algèbre.
- Les 3 langages sont donc équivalents d’un point de vue puissance d’expression.
- L’algèbre est un langage procédurale (quoi et comment) alors que le calcul ne l’est pas (seulement quoi)